Sampling algorithms for generalized model ensembles in multifidelity uncertainty quantification

Workshop on Multilevel and multifidelity samplng methods in UQ for PDEs, Erwin Schrödinger Institute (Virtual)
May 4-5, 2020

Alex Gorodetsky
joint work with Gianluca Geraci, Mike Eldred and John Jakeman (SNL)

UNIVERSITY OF MICHIGAN

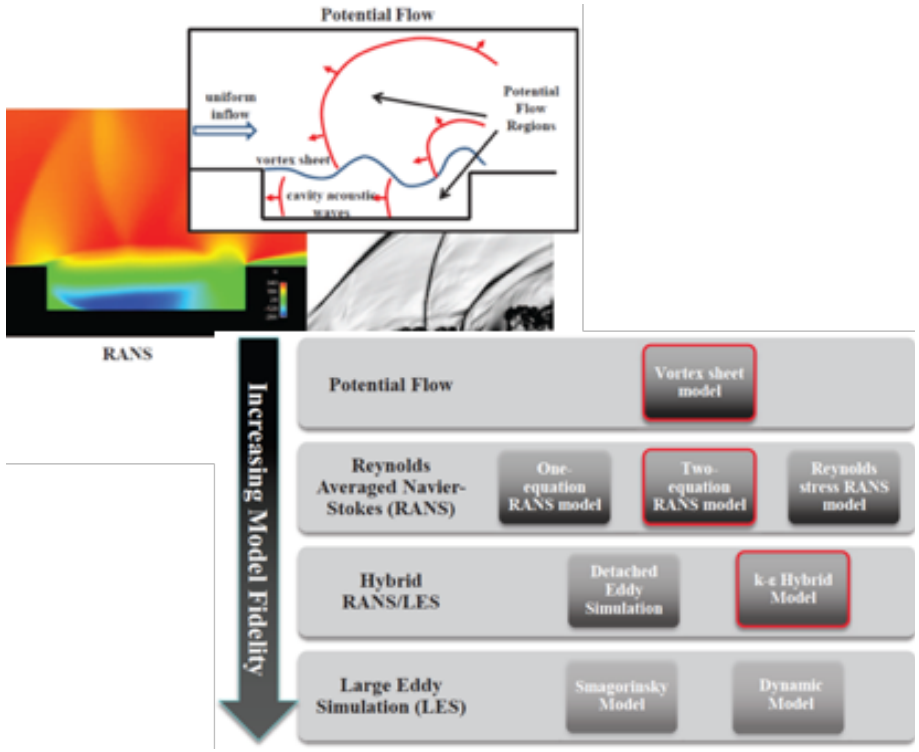
Uncertainty quantification with multiple models

Motivation

Observation: Many models and simulations are created over the course of analyzing a system.

Key idea: Usually, these models serve to **develop** the ***eventual*** high-fidelity model. How do we use them to also **analyze** the eventual high-fidelity model?.

- Information sources
  - Hierarchy of fidelities
  - Ensemble of peer models
  - Discretization levels
  - Experimental data



High-level modeling questions

1. How do we model the multi-fidelity uncertainty quantification problem?
   - How is each simulation related? (Discretization hierarchies, physics hierarchies, "competing" peers (model selection), combination)
   - How are relationships between models represented/exploited? (discrepancies, statistical correlations, probabilistic conditional dependencies, something else?)

High-level modeling questions

1. How do we model the multi-fidelity uncertainty quantification problem?
   - How is each simulation related? (Discretization hierarchies, physics hierarchies, "competing" peers (model selection), combination)
   - How are relationships between models represented/exploited? (discrepancies, statistical correlations, probabilistic conditional dependencies, something else?)
2. What are the algorithmic objectives of multi-fidelity uncertainty quantification?
   - Variance/Bias reduction, MSE reduction (Frequentist)
   - Probabilistic predictions (Bayesian)

High-level modeling questions

1. How do we model the multi-fidelity uncertainty quantification problem?
   - How is each simulation related? (Discretization hierarchies, physics hierarchies, "competing" peers (model selection), combination)
   - How are relationships between models represented/exploited? (discrepancies, statistical correlations, probabilistic conditional dependencies, something else?)
2. What are the algorithmic objectives of multi-fidelity uncertainty quantification?
   - Variance/Bias reduction, MSE reduction (Frequentist)
   - Probabilistic predictions (Bayesian)
3. Are there any *general* approaches for addressing these modeling questions, which can then be adapted to specific cases?

Algorithmic challenges

1. Often a database of evaluations is already available (legacy data), how do we use it without further queries?

Algorithmic challenges

UNIVERSITY OF MICHIGAN

1. Often a database of evaluations is already available (legacy data), how do we use it without further queries?
2. Data has "noise" stemming from different sources
   - Corrupted evaluations
   - Unconverged grids, because simulations tuned to fixed parameter settings
   - Problem geometry leads to meshing issues

Algorithmic challenges

UNIVERSITY OF MICHIGAN

1. Often a database of evaluations is already available (legacy data), how do we use it without further queries?
2. Data has "noise" stemming from different sources
   - Corrupted evaluations
   - Unconverged grids, because simulations tuned to fixed parameter settings
   - Problem geometry leads to meshing issues
3. Nonlinearities in time-dependent problems (dynamical systems)
   - Correlations may not be the best "structure" to exploit here
   - How can chaotic systems be considered?

Algorithmic challenges

UNIVERSITY OF MICHIGAN

1. Often a database of evaluations is already available (legacy data), how do we use it without further queries?
2. Data has "noise" stemming from different sources
   - Corrupted evaluations
   - Unconverged grids, because simulations tuned to fixed parameter settings
   - Problem geometry leads to meshing issues
3. Nonlinearities in time-dependent problems (dynamical systems)
   - Correlations may not be the best "structure" to exploit here
   - How can chaotic systems be considered?
4. Simulation models have different inputs
   - Extremely common in variety of physics domain where lower fidelities replace "physics" with "models"
   - Each low-fidelity may have its own set of uncertain model parameters, while all share uncertain environmental parameters (e.g., boundary conditions)

This talk

Dealing with general model ensembles

We present several ***sampling*** approaches that avoid explicit or implicit orderings based on model fidelity or correlations.
We focus on ***variance reduction***, with some broader ideas at the end

1. Take an ***optimality and convergence*** viewpoint for variance reduction: What is the best possible performance when increasing the information from each low-fidelity information source?

[Gorodetsky, et. al. *A generalized approximate control variate framework for multifidelity uncertainty quantification*, JCP 2020]

[Gorodetsky, et. al. *MFNETS: multi-fidelity data-driven networks for Bayesian learning and prediction*, (submitted) 2020]

This talk

Dealing with general model ensembles

We present several ***sampling*** approaches that avoid explicit or implicit orderings based on model fidelity or correlations.
We focus on ***variance reduction***, with some broader ideas at the end

1. Take an ***optimality and convergence*** viewpoint for variance reduction: What is the best possible performance when increasing the information from each low-fidelity information source?
2. Some commonly used existing approaches built on recursive-difference and recursive-nested estimators have limited variance reduction
   - Low-fidelity sims can be infinitely evaluated with no further variance reduction
   - Arises because of orderings and recursive sampling strategies

[Gorodetsky, et. al. *A generalized approximate control variate framework for multifidelity uncertainty quantification*, JCP 2020]

[Gorodetsky, et. al. *MFNETS: multi-fidelity data-driven networks for Bayesian learning and prediction*, (submitted) 2020]

This talk

Dealing with general model ensembles

We present several ***sampling*** approaches that avoid explicit or implicit orderings based on model fidelity or correlations.
We focus on ***variance reduction***, with some broader ideas at the end

1. Take an ***optimality and convergence*** viewpoint for variance reduction: What is the best possible performance when increasing the information from each low-fidelity information source?
2. Some commonly used existing approaches built on recursive-difference and recursive-nested estimators have limited variance reduction
   - Low-fidelity sims can be infinitely evaluated with no further variance reduction
   - Arises because of orderings and recursive sampling strategies
3. Convergent estimators can be derived

[Gorodetsky, et. al. *A generalized approximate control variate framework for multifidelity uncertainty quantification*, JCP 2020]

[Gorodetsky, et. al. *MFNETS: multi-fidelity data-driven networks for Bayesian learning and prediction*, (submitted) 2020]

Outline

1. Notation: Monte Carlo and control variates
2. Recursive difference (MLMC) and nested difference estimators (MFMC)
3. Convergent approximate control variates
4. Probabilistic models

Monte Carlo

1. z: input parameter uncertainties
2. Q: quantity of interest evaluation

$$\hat{Q}(\boldsymbol{z}) = \frac{1}{N}\sum_{i=1}^{N} Q(z^{(i)})$$

3. The estimator is unbiased

$$\mathbb{E}\left[\hat{Q}(\boldsymbol{z})\right] = \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}\left[Q(z^{(i)})\right] = \mathbb{E}\left[Q\right]$$

4. The estimator variance decays like $1/N$:

$$\mathbb{V}\text{ar}[\hat{Q}] = \frac{1}{N}\sum_{i=1}^{N} \mathbb{V}\text{ar}\left[Q(z^{(i)})\right] = \frac{\mathbb{V}\text{ar}[Q]}{N}$$

Linear control variate
Single additional information source

- Introduce a correlated information source $\hat{Q}_1(\boldsymbol{z})$

[Lavenberg and Welch 1981]

Linear control variate
Single additional information source

- Introduce a correlated information source $\hat{Q}_1(\boldsymbol{z})$
- Suppose we **know** the mean of $\mathbb{E}\left[\hat{Q}_1(\boldsymbol{z})\right] = \mu_1$

$$\hat{Q}^{\text{CV}} = \hat{Q}(\boldsymbol{z}) + \alpha(\hat{Q}_1(\boldsymbol{z}) - \mu_1)$$

[Lavenberg and Welch 1981]

Linear control variate
Single additional information source

- Introduce a correlated information source $\hat{Q}_1(\boldsymbol{z})$
- Suppose we **know** the mean of $\mathbb{E}\left[\hat{Q}_1(\boldsymbol{z})\right] = \mu_1$

$$\hat{Q}^{\text{CV}} = \hat{Q}(\boldsymbol{z}) + \alpha(\hat{Q}_1(\boldsymbol{z}) - \mu_1)$$

- Variance reduction effectiveness measured by correlation $\rho_1(\hat{Q}(\boldsymbol{z}), \hat{Q}_1(\boldsymbol{z}))$

$$\mathbb{V}\text{ar}[\hat{Q}^{\text{CV}}] = (1 - \rho_1^2)\mathbb{V}\text{ar}[\hat{Q}]$$

[Lavenberg and Welch 1981]

Linear control variate
Multiple Models

- Several additional random variables $\hat{Q}_i(\boldsymbol{z})$ with known means μ_i

$$\hat{Q}^{\text{CV}} = \hat{Q}(\boldsymbol{z}) + \sum_{i=1}^{M} \alpha_i (\hat{Q}_i(\boldsymbol{z}) - \mu_i)$$

[Kleijnen 1974, Iglehart 1979, Lavenberg 1982, Rubinstein 1985]

Linear control variate
Multiple Models

- Several additional random variables $\hat{Q}_i(\boldsymbol{z})$ with known means μ_i

$$\hat{Q}^{\mathsf{CV}} = \hat{Q}(\boldsymbol{z}) + \sum_{i=1}^{M} \alpha_i (\hat{Q}_i(\boldsymbol{z}) - \mu_i)$$

$$\mathbb{V}\mathrm{ar}[\hat{Q}^{\mathsf{CV}}] = (1 - R^2)\mathbb{V}\mathrm{ar}[\hat{Q}]$$

[Kleijnen 1974, Iglehart 1979, Lavenberg 1982, Rubinstein 1985]

Linear control variate
Multiple Models

- Several additional random variables $\hat{Q}_i(\boldsymbol{z})$ with known means μ_i

$$\hat{Q}^{\text{CV}} = \hat{Q}(\boldsymbol{z}) + \sum_{i=1}^{M} \alpha_i (\hat{Q}_i(\boldsymbol{z}) - \mu_i)$$

$$\mathbb{V}\text{ar}[\hat{Q}^{\text{CV}}] = (1 - R^2)\mathbb{V}\text{ar}[\hat{Q}]$$

- Optimal weights are

$$\alpha = -\mathbb{C}\text{ov}[\hat{Q}, \hat{\boldsymbol{Q}}]\mathbb{C}\text{ov}[\hat{\boldsymbol{Q}}, \hat{\boldsymbol{Q}}]^{-1}$$

- Optimal reduction is (**note covariance amongst all low-fidelity sources**)

$$R^2 = \frac{\mathbb{C}\text{ov}[\hat{Q}, \hat{\boldsymbol{Q}}]\mathbb{C}\text{ov}[\hat{\boldsymbol{Q}}, \hat{\boldsymbol{Q}}]^{-1}\mathbb{C}\text{ov}[\hat{\boldsymbol{Q}}, \hat{Q}]}{\mathbb{V}\text{ar}[\hat{Q}]}$$

[Kleijnen 1974, Iglehart 1979, Lavenberg 1982, Rubinstein 1985]

Adaptation to UQ

- Additional information comes from other simulations and/or data

Adaptation to UQ

- Additional information comes from other simulations and/or data
- The optimal linear control variate has components we do not have

$$\hat{Q}^{\text{CV}} = \hat{Q}(\boldsymbol{z}) + \sum_{i=1}^{M} \alpha_i (\hat{Q}_i(\boldsymbol{z}) - \mu_i)$$

Adaptation to UQ

- Additional information comes from other simulations and/or data
- The optimal linear control variate has components we do not have

$$\hat{Q}^{\text{CV}} = \hat{Q}(\boldsymbol{z}) + \sum_{i=1}^{M} \alpha_i (\hat{Q}_i(\boldsymbol{z}) - \mu_i)$$

- Approximate the control variate by generating estimators as (**heuristic model of MFUQ**)

$$\tilde{Q}(\boldsymbol{z}, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_M) = \hat{Q}(\boldsymbol{z}) + \sum_{i=1}^{M} \alpha_i (\hat{Q}_i(\boldsymbol{z}_i^1) - \hat{\mu}_i(\boldsymbol{z}_i^2))$$

Adaptation to UQ

- Additional information comes from other simulations and/or data
- The optimal linear control variate has components we do not have

$$\hat{Q}^{\text{CV}} = \hat{Q}(\boldsymbol{z}) + \sum_{i=1}^{M} \alpha_i (\hat{Q}_i(\boldsymbol{z}) - \mu_i)$$

- Approximate the control variate by generating estimators as (**heuristic model of MFUQ**)

$$\tilde{Q}(\boldsymbol{z}, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_M) = \hat{Q}(\boldsymbol{z}) + \sum_{i=1}^{M} \alpha_i (\hat{Q}_i(\boldsymbol{z}_i^1) - \hat{\mu}_i(\boldsymbol{z}_i^2))$$

- Single control variate case considered by [Pasupathy, Schmeiser, Taaffee, and Wang 2012] and [Ng and Willcox, 2014]

Multiple-model case

- Two families of recursive algorithms can be understood with this model
  - Recursive difference estimator (inner loop of MLMC and MIMC, [Giles, 2008], [Haji-Ali, Nobile, Tempone 2016])
  - Recursive nested estimator (Multifidelity Monte Carlo [Peherstorfer, Willcox, and Gunzberger 2016])
- Variance reduction of these recursive approaches do not converge to optimal control variate in the limit of infinite data
- Next: an interlude about robustness of some recursive approaches
- Eventual: Convergent and robust ACV estimators can be found

A limitation of some recursive schemes
Robustness to noise

- Stochastic ODE with payoff modeling European options (Giles 2008)

$$dS = rSdt + \sigma SdW \quad P = \exp(-r)\max(0, S(1) - 1)$$

- Euler-Maruyama for time integration at **different resolutions**
- Expectation is a telescopic sum —

$$\mathbb{E}[P_L] = \mathbb{E}[P_1] + \sum_{\ell=2}^{L} \mathbb{E}[P_i - P_{i-1}] \quad \mathbb{V}\text{ar}[P_L] = \mathbb{V}\text{ar}[P_1] + \sum_{\ell=2}^{L} \mathbb{V}\text{ar}[P_i - P_{i-1}]$$

A limitation of some recursive schemes
Robustness to noise

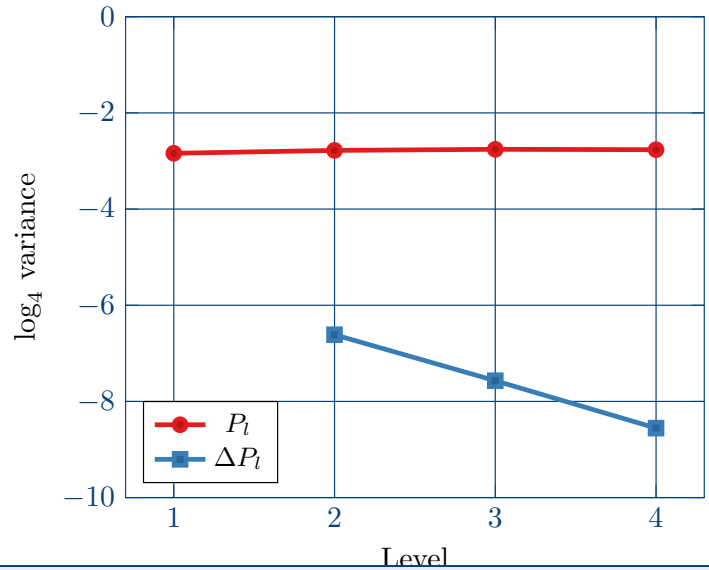
- Stochastic ODE with payoff modeling European options (Giles 2008)

$$dS = rSdt + \sigma SdW \quad P = \exp(-r)\max(0, S(1) - 1)$$

- Euler-Maruyama for time integration at **different resolutions**
- Expectation is a telescopic sum —

$$\mathbb{E}[P_L] = \mathbb{E}[P_1] + \sum_{\ell=2}^{L} \mathbb{E}[P_i - P_{i-1}] \quad \mathbb{V}\text{ar}[P_L] = \mathbb{V}\text{ar}[P_1] + \sum_{\ell=2}^{L} \mathbb{V}\text{ar}[P_i - P_{i-1}]$$

Variance Decay → MLMC



A limitation of some recursive schemes
Robustness to noise

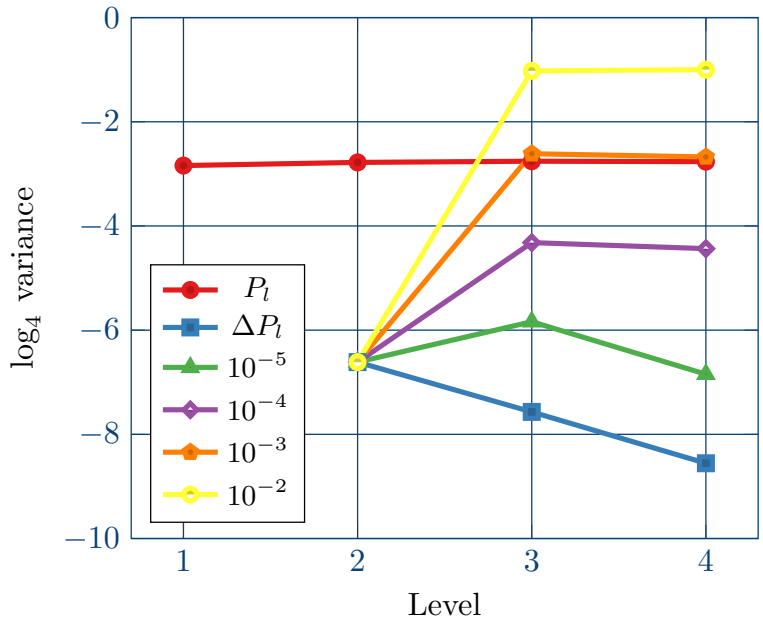
- Stochastic ODE with payoff modeling European options (Giles 2008)

$$dS = rSdt + \sigma SdW \quad P = \exp(-r)\max(0, S(1) - 1)$$

- Euler-Maruyama for time integration at **different resolutions**
- Expectation is a telescopic sum —

$$\mathbb{E}[P_L] = \mathbb{E}[P_1] + \sum_{\ell=2}^{L} \mathbb{E}[P_i - P_{i-1}] \quad \mathbb{V}\text{ar}[P_L] = \mathbb{V}\text{ar}[P_1] + \sum_{\ell=2}^{L} \mathbb{V}\text{ar}[P_i - P_{i-1}]$$

Corrupt some level 3 evals.

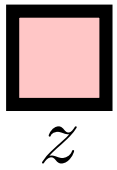


Recursive differences ***within*** Multilevel Monte Carlo

1. Consider a sequence of LF sources

$$\{Q_1, \ldots, Q_M\}$$

2. Recursive difference est. [Owen 2013]



$$\tilde{Q} = (Q - Q_1) + \mu_1 \text{ so that } \mathbb{E}\left[\tilde{Q}\right] = \mathbb{E}\left[Q\right]$$

Recursive differences ***within*** Multilevel Monte Carlo

1. Consider a sequence of LF sources

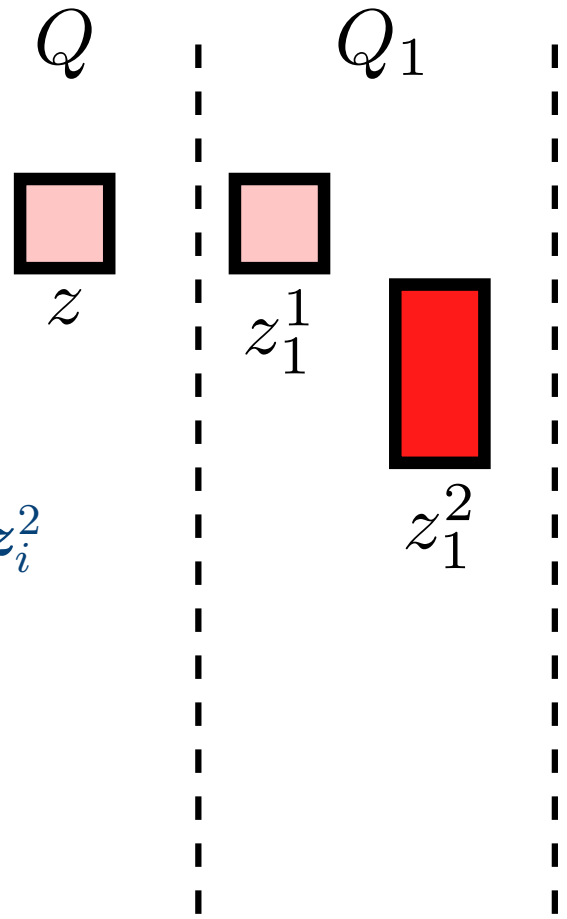
$$\{Q_1, \ldots, Q_M\}$$

2. Recursive difference est. [Owen 2013]

$$\tilde{Q} = (Q - Q_1) + \mu_1 \text{ so that } \mathbb{E}\left[\tilde{Q}\right] = \mathbb{E}\left[Q\right]$$

3. Samples for each level partioned $\boldsymbol{z}_i = \boldsymbol{z}_i^1 \cup \boldsymbol{z}_i^2$

$$\hat{Q}^{\text{MLMC}}(\boldsymbol{z}) = \left(\hat{Q}(\boldsymbol{z}) - \hat{Q}_1(\boldsymbol{z}_1^1)\right) + \hat{\mu}_1(\boldsymbol{z}_1^2)$$



Recursive differences ***within*** Multilevel Monte Carlo

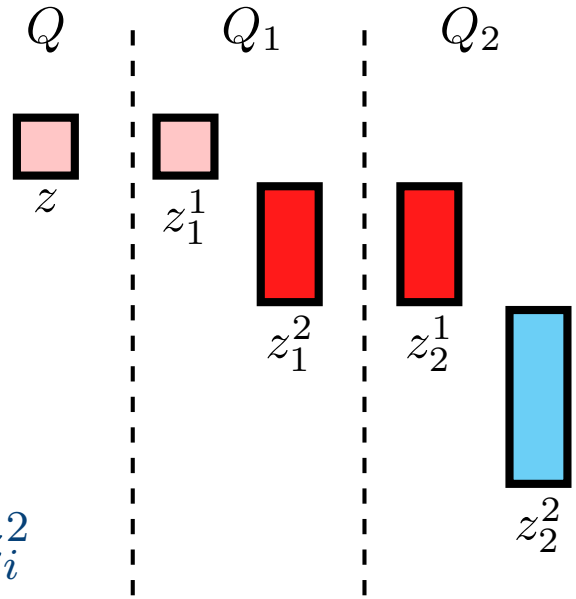
1. Consider a sequence of LF sources

$$\{Q_1, \ldots, Q_M\}$$

2. Recursive difference est. [Owen 2013]

$$\tilde{Q} = (Q - Q_1) + \mu_1 \text{ so that } \mathbb{E}\left[\tilde{Q}\right] = \mathbb{E}\left[Q\right]$$

3. Samples for each level partioned $\boldsymbol{z}_i = \boldsymbol{z}_i^1 \cup \boldsymbol{z}_i^2$



$$\hat{Q}^{\text{MLMC}}(\boldsymbol{z}) = \left(\hat{Q}(\boldsymbol{z}) - \hat{Q}_1(\boldsymbol{z}_1^1)\right) + \left(\hat{\mu}_1(\boldsymbol{z}_1^2) - \hat{Q}_2(\boldsymbol{z}_1^2)\right) + \hat{\mu}_2(\boldsymbol{z}_2^2)$$

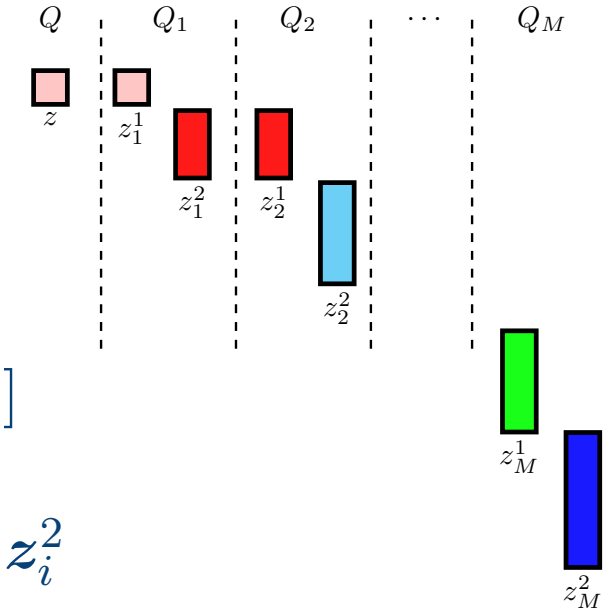
Recursive differences ***within*** Multilevel Monte Carlo

1. Consider a sequence of LF sources

$$\{Q_1, \ldots, Q_M\}$$

2. Recursive difference est. [Owen 2013]

$$\tilde{Q} = (Q - Q_1) + \mu_1 \text{ so that } \mathbb{E}\left[\tilde{Q}\right] = \mathbb{E}\left[Q\right]$$



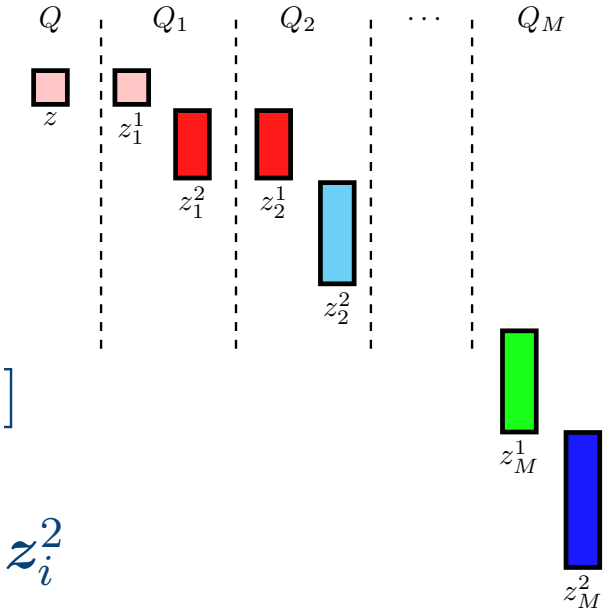
3. Samples for each level partioned $\boldsymbol{z}_i = \boldsymbol{z}_i^1 \cup \boldsymbol{z}_i^2$

$$\hat{Q}^{\text{MLMC}}(\boldsymbol{z}) = \left(\hat{Q}(\boldsymbol{z}) - \hat{Q}_1(\boldsymbol{z}_1^1)\right) + \left(\hat{\mu}_1(\boldsymbol{z}_1^2) - \hat{Q}_2(\boldsymbol{z}_1^2)\right) + \hat{\mu}_2(\boldsymbol{z}_2^2) + \ldots$$

Recursive differences ***within*** Multilevel Monte Carlo

1. Consider a sequence of LF sources

$$\{Q_1, \ldots, Q_M\}$$



2. Recursive difference est. [Owen 2013]

$$\tilde{Q} = (Q - Q_1) + \mu_1 \text{ so that } \mathbb{E}\left[\tilde{Q}\right] = \mathbb{E}\left[Q\right]$$

3. Samples for each level partioned $\boldsymbol{z}_i = \boldsymbol{z}_i^1 \cup \boldsymbol{z}_i^2$

$$\hat{Q}^{\text{MLMC}}(\boldsymbol{z}) = \left(\hat{Q}(\boldsymbol{z}) - \hat{Q}_1(\boldsymbol{z}_1^1)\right) + \left(\hat{\mu}_1(\boldsymbol{z}_1^2) - \hat{Q}_2(\boldsymbol{z}_1^2)\right) + \hat{\mu}_2(\boldsymbol{z}_2^2) + \ldots$$

4. This is an ACV with $\alpha = -1$ and variance

$$\mathbb{V}\text{ar}[Q] = \mathbb{V}\text{ar}[Q_L] + \sum_{\ell=L-1}^{1} \mathbb{V}\text{ar}[Q_\ell - Q_{\ell+1}] + \mathbb{V}\text{ar}[Q - Q_1]$$

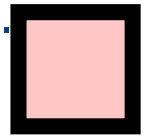
Recursive-nested estimators ***within*** Multifidelity Monte Carlo

1. Consider a sequence of LF sources

$$\{Q_1, \ldots, Q_M\}$$

Q

2. Recursive nested estimator, base is the CV ..



z

$$\tilde{Q} = \hat{Q} + \alpha \left(\hat{Q}_1 - \mu_1 \right)$$

$$\mathbb{E}\left[\tilde{Q}\right] = \mathbb{E}\left[Q\right]$$

Recursive-nested estimators ***within*** Multifidelity Monte Carlo

1. Consider a sequence of LF sources

$$\{Q_1, \ldots, Q_M\}$$

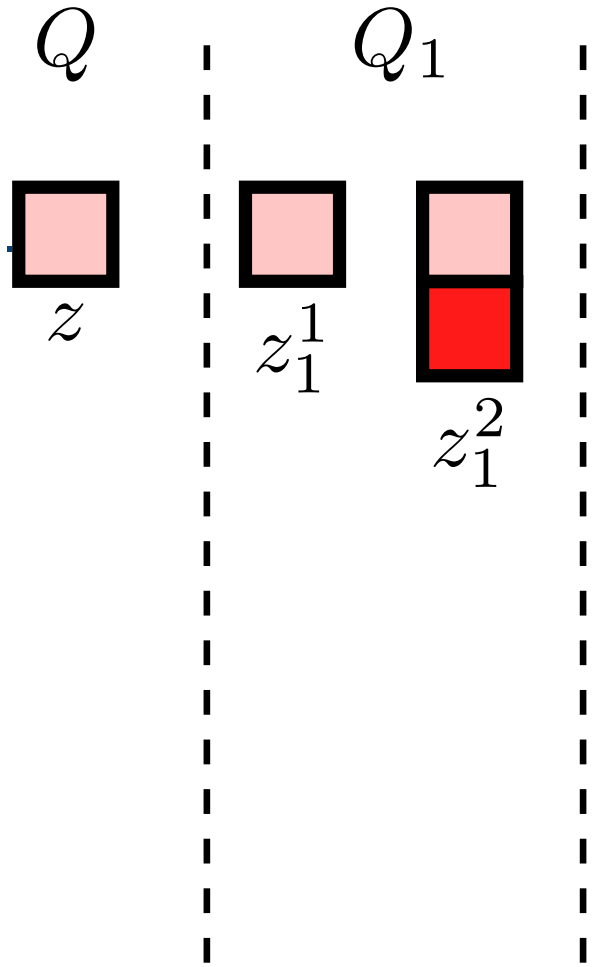
2. Recursive nested estimator, base is the CV ...

$$\tilde{Q} = \hat{Q} + \alpha \left(\hat{Q}_1 - \mu_1 \right)$$

$$\mathbb{E}\left[\tilde{Q}\right] = \mathbb{E}\left[Q\right]$$

3. ... however samples are nested

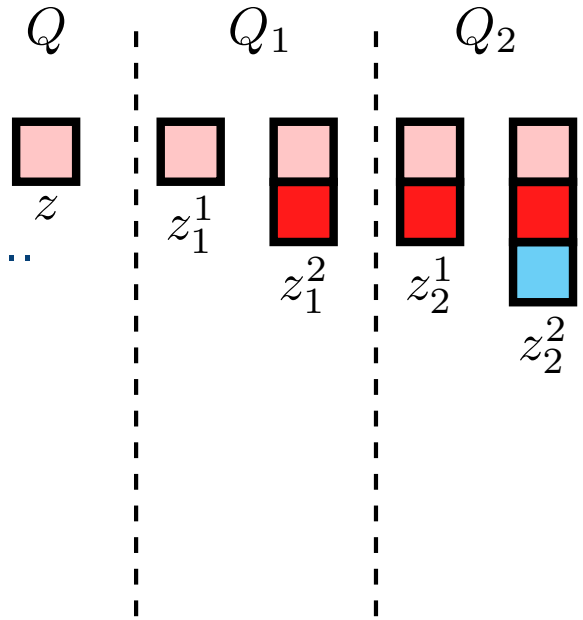
$$\hat{Q}^{\text{MFMC}}(\boldsymbol{z}) = \hat{Q}(\boldsymbol{z}) + \alpha_1(\hat{Q}_1(\boldsymbol{z}_1^1) - \hat{\mu}_1(\boldsymbol{z}_1^2))$$



Recursive-nested estimators ***within*** Multifidelity Monte Carlo

1. Consider a sequence of LF sources

$$\{Q_1, \ldots, Q_M\}$$



2. Recursive nested estimator, base is the CV ...

$$\tilde{Q} = \hat{Q} + \alpha \left(\hat{Q}_1 - \mu_1 \right)$$

$$\mathbb{E}\left[\tilde{Q}\right] = \mathbb{E}\left[Q\right]$$

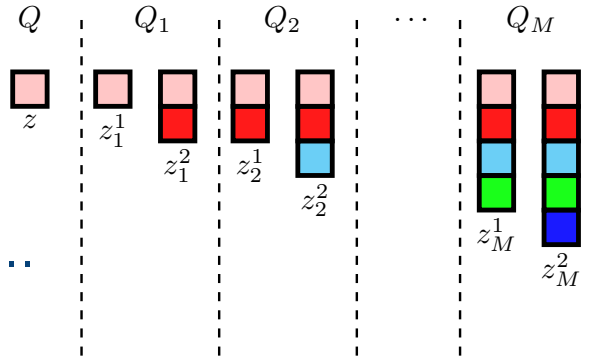
3. ... however samples are nested

$$\hat{Q}^{\text{MFMC}}(\boldsymbol{z}) = \hat{Q}(\boldsymbol{z}) + \alpha_1 (\hat{Q}_1(\boldsymbol{z}_1^1) - \left(\hat{\mu}_1(\boldsymbol{z}_1^2) + \alpha_2 \left(\hat{Q}_2(\boldsymbol{z}_1^2) - \hat{\mu}_2(\boldsymbol{z}_2^2) \right) \right)$$

Recursive-nested estimators ***within*** Multifidelity Monte Carlo

1. Consider a sequence of LF sources

$$\{Q_1, \ldots, Q_M\}$$



2. Recursive nested estimator, base is the CV ...

$$\tilde{Q} = \hat{Q} + \alpha \left(\hat{Q}_1 - \mu_1 \right)$$

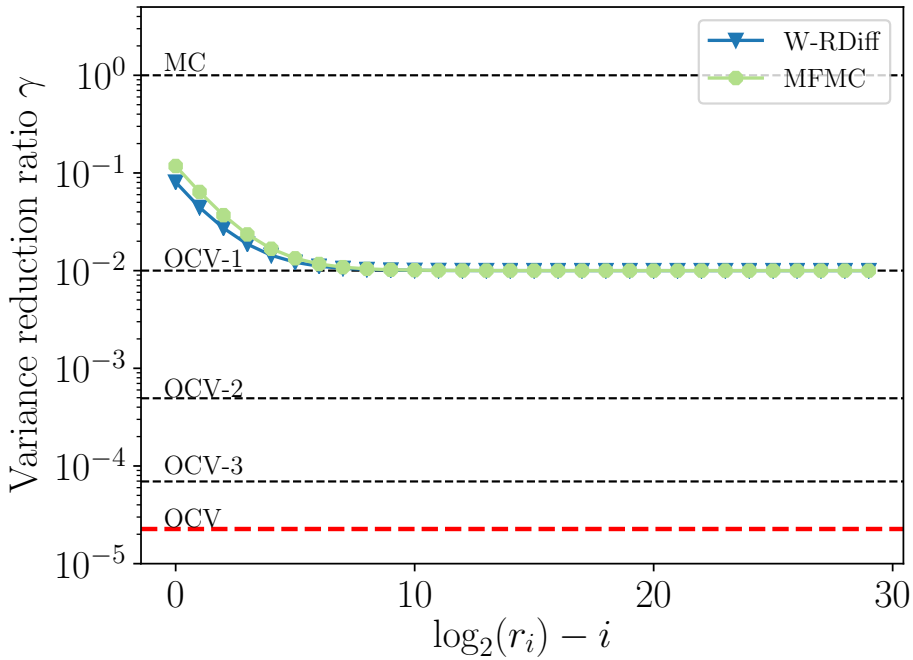
$$\mathbb{E}\left[\tilde{Q}\right] = \mathbb{E}\left[Q\right]$$

3. ... however samples are nested

$$\hat{Q}^{\text{MFMC}}(\boldsymbol{z}) = \hat{Q}(\boldsymbol{z}) + \alpha_1 (\hat{Q}_1(\boldsymbol{z}_1^1) - \left(\hat{\mu}_1(\boldsymbol{z}_1^2) + \alpha_2 \left(\hat{Q}_2(\boldsymbol{z}_1^2) - (\hat{\mu}_2(\boldsymbol{z}_2^2) + \ldots) \right) \right)$$

Potentially significant limitations to variance reduction

- **Fix** number of HF evaluations
- Increase the number of LF evaluations
- $Q = x^5$, $Q_1 = x^4$, $Q_2 = x^3$, $Q_3 = x^2$, $Q_4 = x$



Sub-optimality of recursive estimators

THEOREM 2.4 (Maximum variance reduction of MLMC). *The variance reduction of MLMC is bounded above by the optimal single CV i.e.,*

$$R^2_{\text{MLMC}} < \rho_1^2. \tag{2.19}$$

THEOREM 2.7 (Maximum variance reduction of MFMC). *The variance reduction of MFMC is bounded above by the optimal single CV, i.e.,*

$$R^2_{\text{MFMC}} < \rho_1^2. \tag{2.25}$$

Sub-optimality of recursive estimators

THEOREM 2.4 (Maximum variance reduction of MLMC). *The variance reduction of MLMC is bounded above by the optimal single CV i.e.,*

$$R^2_{\text{MLMC}} < \rho_1^2. \tag{2.19}$$

THEOREM 2.7 (Maximum variance reduction of MFMC). *The variance reduction of MFMC is bounded above by the optimal single CV, i.e.,*

$$R^2_{\text{MFMC}} < \rho_1^2. \tag{2.25}$$

Why?: lack of consideration of all correlations $R^2 = \frac{\mathbb{C}\text{ov}[\hat{Q},\hat{\boldsymbol{Q}}]\mathbb{C}\text{ov}[\hat{\boldsymbol{Q}},\hat{\boldsymbol{Q}}]^{-1}\mathbb{C}\text{ov}[\hat{\boldsymbol{Q}},\hat{Q}]}{\mathbb{V}\text{ar}[\hat{Q}]}$

Sub-optimality of recursive estimators

THEOREM 2.4 (Maximum variance reduction of MLMC). *The variance reduction of MLMC is bounded above by the optimal single CV i.e.,*

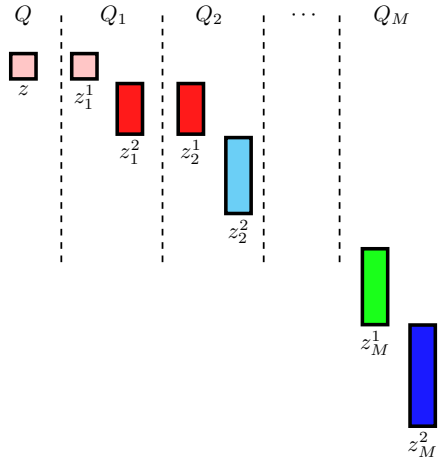
$$R^2_{\text{MLMC}} < \rho_1^2. \tag{2.19}$$

THEOREM 2.7 (Maximum variance reduction of MFMC). *The variance reduction of MFMC is bounded above by the optimal single CV, i.e.,*

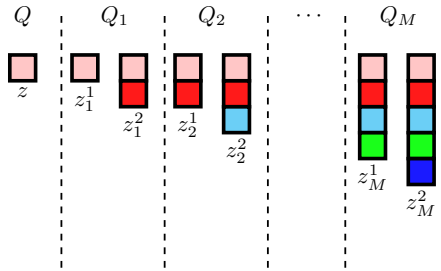
$$R^2_{\text{MFMC}} < \rho_1^2. \tag{2.25}$$

Why?: lack of consideration of all correlations $R^2 = \frac{\mathbb{C}\text{ov}[\hat{Q},\hat{\boldsymbol{Q}}]\mathbb{C}\text{ov}[\hat{\boldsymbol{Q}},\hat{\boldsymbol{Q}}]^{-1}\mathbb{C}\text{ov}[\hat{\boldsymbol{Q}},\hat{Q}]}{\mathbb{V}\text{ar}[\hat{Q}]}$

Recursive Difference



MFMC



Approximate control variates

Estimator Ansatz

$$\tilde{Q}(\boldsymbol{z}, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_M) = \hat{Q}(\boldsymbol{z}) + \sum_{i=1}^{M} \alpha_i (\hat{Q}_i(\boldsymbol{z}) - \hat{\mu}_i(\boldsymbol{z}_i^2))$$

- Ests. distinguished by distribution of samples amongst $\hat{Q}_i$ and μ_i
- Includes previous recursive estimators as specific cases

Approximate control variates

Estimator Ansatz

$$\tilde{Q}(\boldsymbol{z}, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_M) = \hat{Q}(\boldsymbol{z}) + \sum_{i=1}^{M} \alpha_i (\hat{Q}_i(\boldsymbol{z}) - \hat{\mu}_i(\boldsymbol{z}_i^2))$$

- Ests. distinguished by distribution of samples amongst $\hat{Q}_i$ and μ_i
- Includes previous recursive estimators as specific cases

- We seek ACV estimators that **converge** to $\gamma = 1 - R^2$

Approximate control variates

Estimator Ansatz

$$\tilde{Q}(\boldsymbol{z}, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_M) = \hat{Q}(\boldsymbol{z}) + \sum_{i=1}^{M} \alpha_i (\hat{Q}_i(\boldsymbol{z}) - \hat{\mu}_i(\boldsymbol{z}_i^2))$$

- Ests. distinguished by distribution of samples amongst $\hat{Q}_i$ and μ_i
- Includes previous recursive estimators as specific cases

- We seek ACV estimators that **converge** to $\gamma = 1 - R^2$
- The simplest such scheme is the following
  1. Use the same number of samples as the high fidelity model to compute $\hat{Q}_i$

Approximate control variates

Estimator Ansatz

$$\tilde{Q}(\boldsymbol{z}, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_M) = \hat{Q}(\boldsymbol{z}) + \sum_{i=1}^{M} \alpha_i (\hat{Q}_i(\boldsymbol{z}) - \hat{\mu}_i(\boldsymbol{z}_i^2))$$

- Ests. distinguished by distribution of samples amongst $\hat{Q}_i$ and μ_i
- Includes previous recursive estimators as specific cases

- We seek ACV estimators that **converge** to $\gamma = 1 - R^2$
- The simplest such scheme is the following
  1. Use the same number of samples as the high fidelity model to compute $\hat{Q}_i$
  2. Use the same and extra samples to compute $\hat{\mu}_i$

Approximate control variates

Estimator Ansatz

$$\tilde{Q}(\boldsymbol{z}, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_M) = \hat{Q}(\boldsymbol{z}) + \sum_{i=1}^{M} \alpha_i (\hat{Q}_i(\boldsymbol{z}) - \hat{\mu}_i(\boldsymbol{z}_i^2))$$

- Ests. distinguished by distribution of samples amongst $\hat{Q}_i$ and μ_i
- Includes previous recursive estimators as specific cases

- We seek ACV estimators that **converge** to $\gamma = 1 - R^2$
- The simplest such scheme is the following
  1. Use the same number of samples as the high fidelity model to compute $\hat{Q}_i$
  2. Use the same and extra samples to compute $\hat{\mu}_i$
  3. Compute optimal weights (there is a formula)

Approximate control variates

Estimator Ansatz

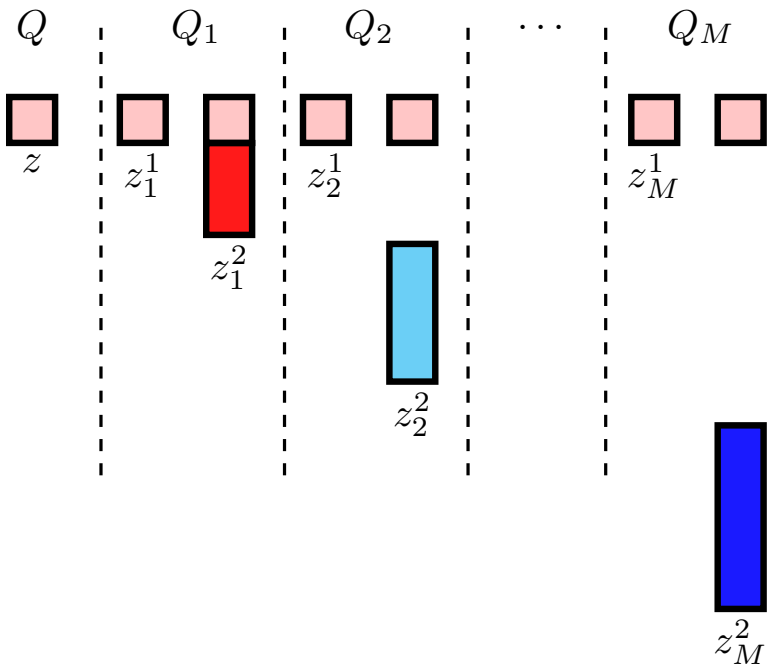
$$\tilde{Q}(\boldsymbol{z}, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_M) = \hat{Q}(\boldsymbol{z}) + \sum_{i=1}^{M} \alpha_i (\hat{Q}_i(\boldsymbol{z}) - \hat{\mu}_i(\boldsymbol{z}_i^2))$$

- Ests. distinguished by distribution of samples amongst $\hat{Q}_i$ and μ_i
- Includes previous recursive estimators as specific cases

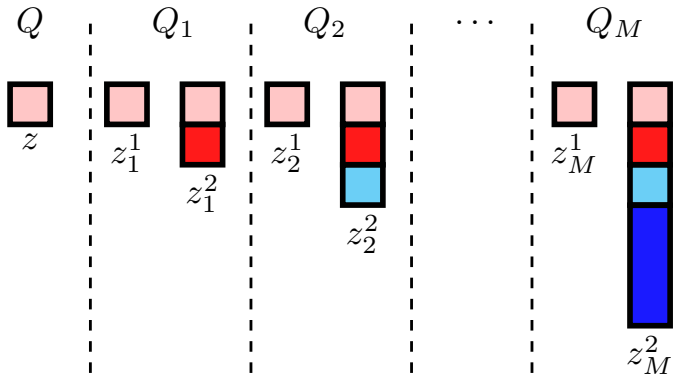
- We seek ACV estimators that **converge** to $\gamma = 1 - R^2$
- The simplest such scheme is the following
  1. Use the same number of samples as the high fidelity model to compute $\hat{Q}_i$
  2. Use the same and extra samples to compute $\hat{\mu}_i$
  3. Compute optimal weights (there is a formula)
- Converges because it ***maintains correlation*** amongst all $\hat{Q}_i$

Sample allocations for two (of many) variants

Independent Samples

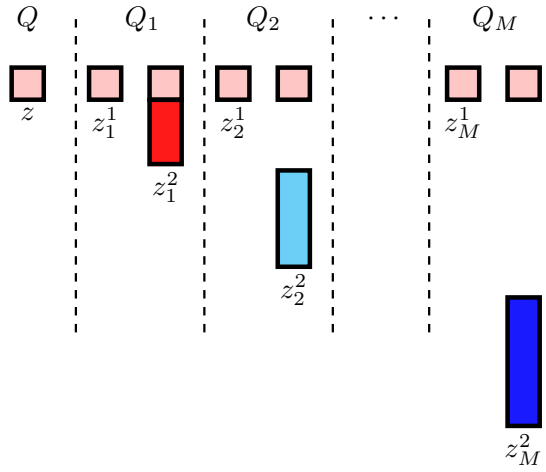


Shared Samples



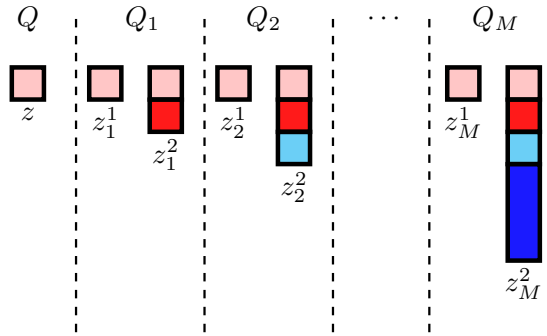
Sample allocations for two (of many) variants

Independent Samples

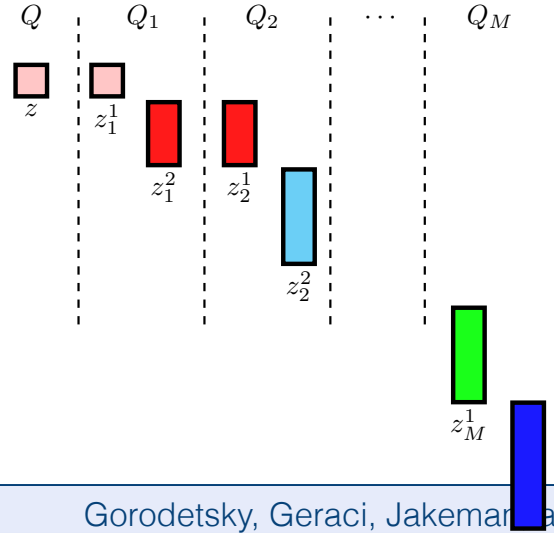


V.S.

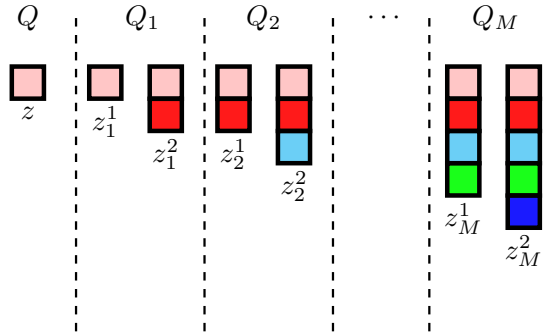
Shared Samples



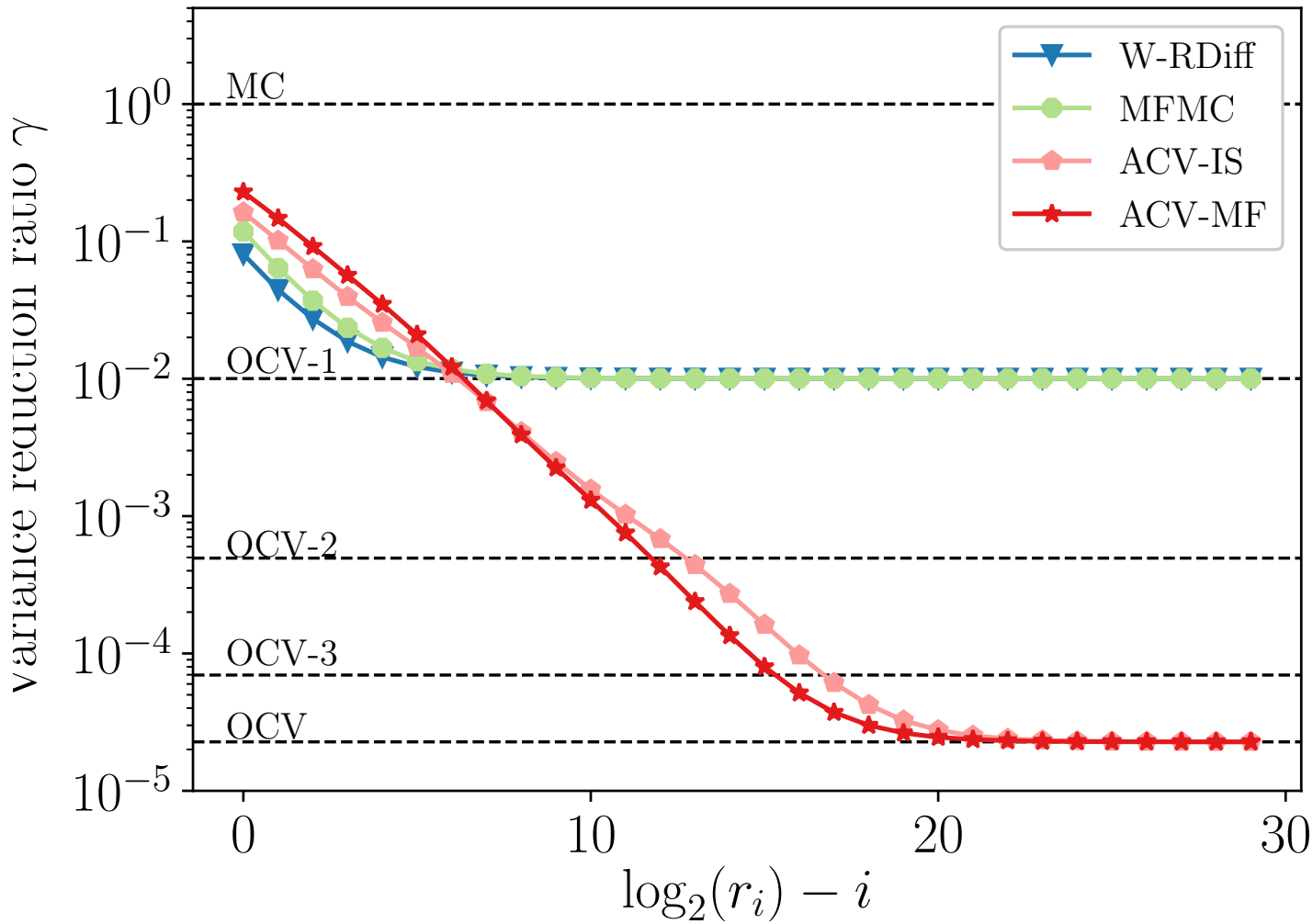
Recursive Difference



MFMC



We can prove convergence



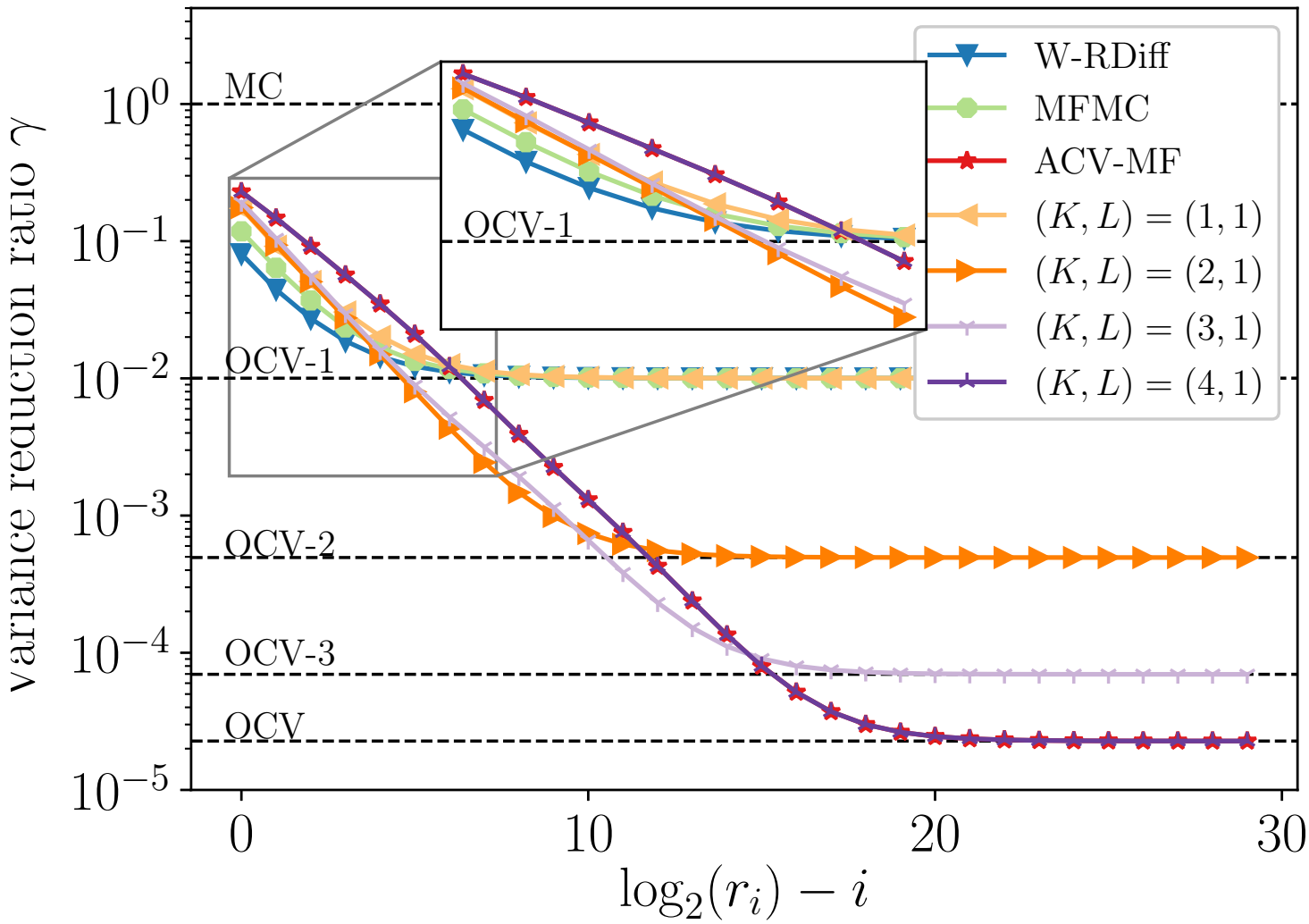
Hybrid estimator to accelerate convergence

- Use the first K control variates to accelerate the Qoi Q
- Use the last $M - K$ to accelerate the convergence of the μ_L for some L.

$$\hat{Q}^{\text{ACV-KL}}(\underline{\alpha}, \underline{\boldsymbol{z}}) = \hat{Q}(\boldsymbol{z}) + \sum_{i=1}^{K} \alpha_i \left(\hat{Q}_i(\boldsymbol{z}) - \hat{\mu}_i(\boldsymbol{z}_i) \right) + \sum_{i=K+1}^{M} \alpha_i \left(\hat{Q}_i(\boldsymbol{z}_L) - \hat{\mu}_i(\boldsymbol{z}_i) \right)$$

$$\hat{Q}^{\text{ACV-KL}}(\underline{\alpha}, \underline{\boldsymbol{z}}) = \hat{Q}_K^{\text{ACV-}MF}(\alpha_1, \ldots, \alpha_K, \boldsymbol{z}, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_k) + \sum_{i=K+1}^{M} \alpha_i \left(\hat{Q}_i(\boldsymbol{z}_L) - \hat{\mu}_i(\boldsymbol{z}_i) \right)$$

Accelerated convergence to target levels



Sample Allocation

1. Optimize over distribution of samples to minimize variance for target cost

$$J_{ACV}(N, \underline{r}) = \left(1 - R^2_{ACV}\right) \frac{\mathbb{C}\text{ov}\left[Q\right]}{N}$$

2. This is a mixed-integer nonlinear program (MINLP) where N is a non-categorical integer variable, $r_i N$ are derived integer quantities, and, in the case of ACV-KL, (K, L) are categorical integer variables.

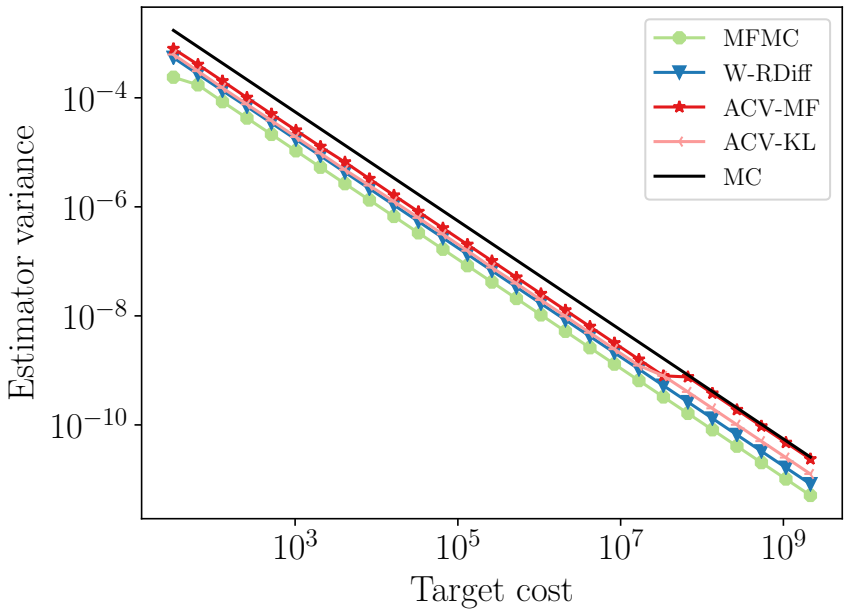
$$\min_{N,\underline{r},K,L} \log(J_{ACV}(N, \underline{r}, K, L)) \quad \text{subject to}$$

$$N\left(w + \sum_{i=1}^{M} w_i r_i\right) \leq C, \quad N \geq 1, \quad r_1 \geq 1, \text{ and } r_i \geq r_{i-1} \text{ for } i = 2, \ldots, M$$

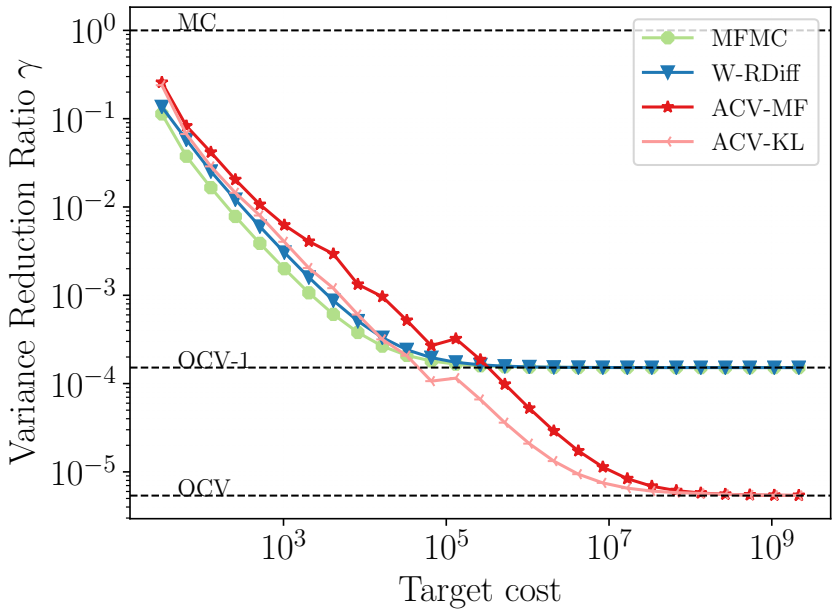
Viscous Burgers with uncertain boundaries

- Compare recursive-difference, MFMC, and optimal ACV
- Viscous Burgers at 6 discretization levels
- Estimate mean of u at a spatial location

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} - \kappa\frac{\partial^2 u}{\partial x^2} = 0$$

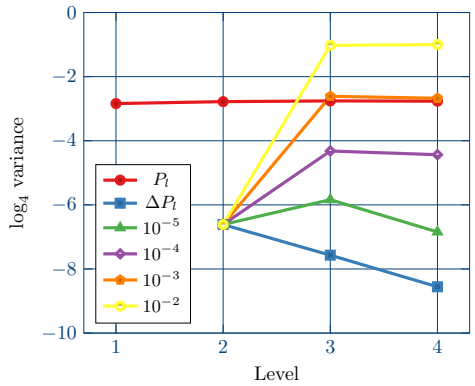


(a) Number of high-fidelity is variable

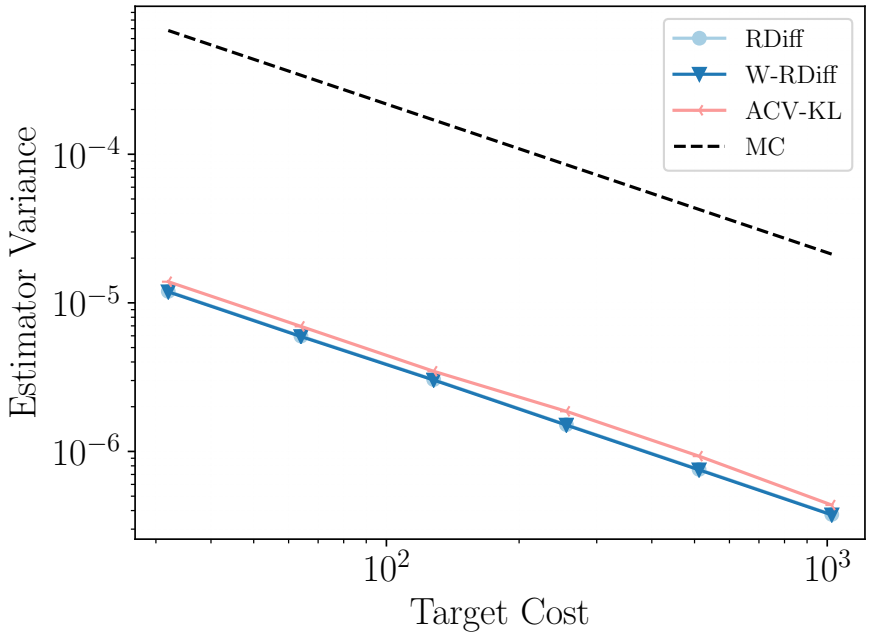


(b) Number of high-fidelity is fixed

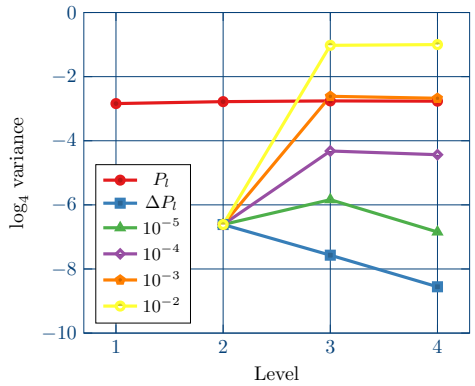
Back to the robustness issue
Considerable improvement by using full correlations



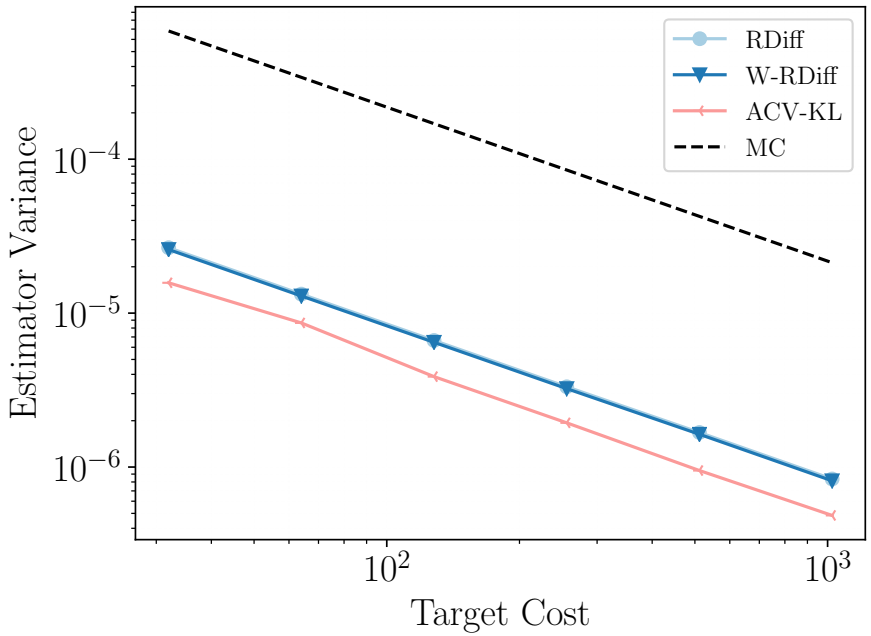
No Corruption



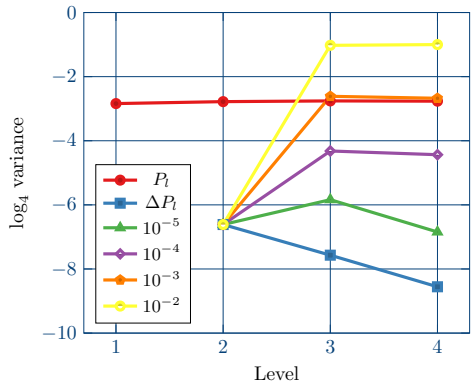
Back to the robustness issue
Considerable improvement by using full correlations



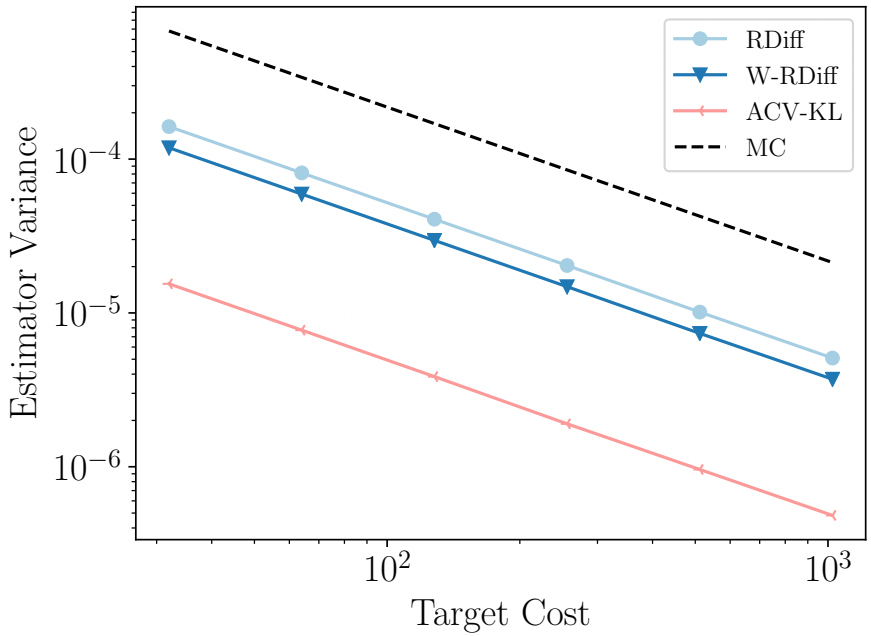
Corruption Rate: 10^{-5}



Back to the robustness issue
Considerable improvement by using full correlations



Corruption Rate: 10^{-4}

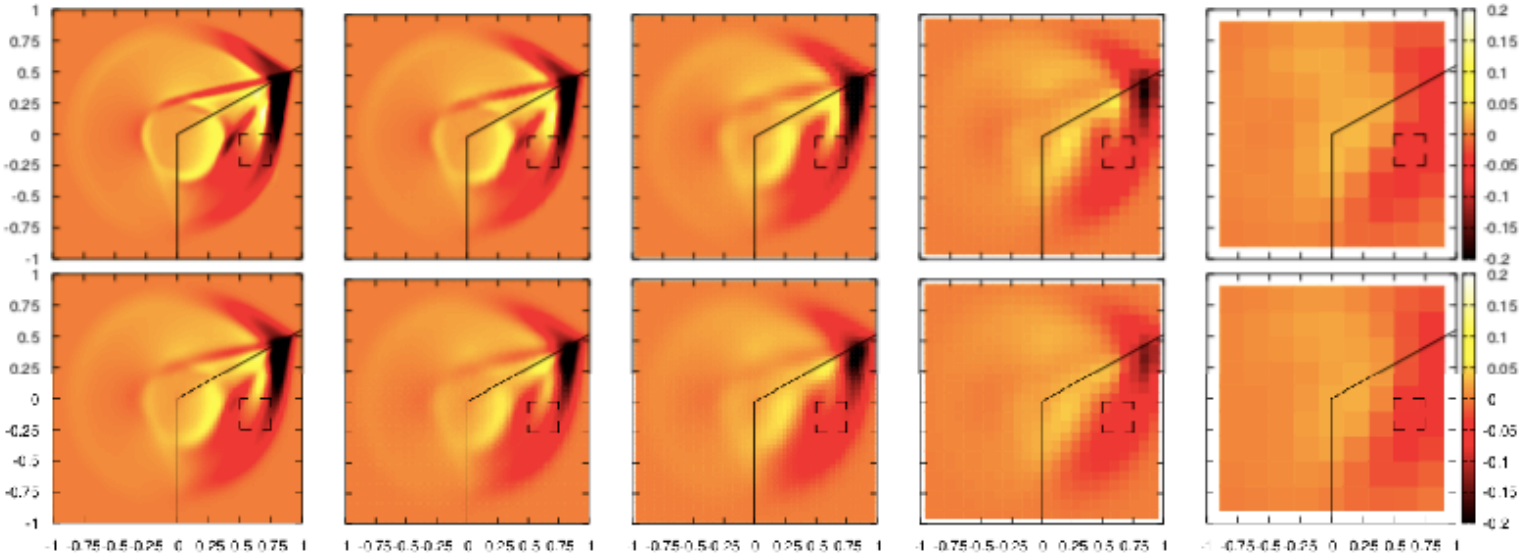


Elastic wave equation

Hyperbolic equations

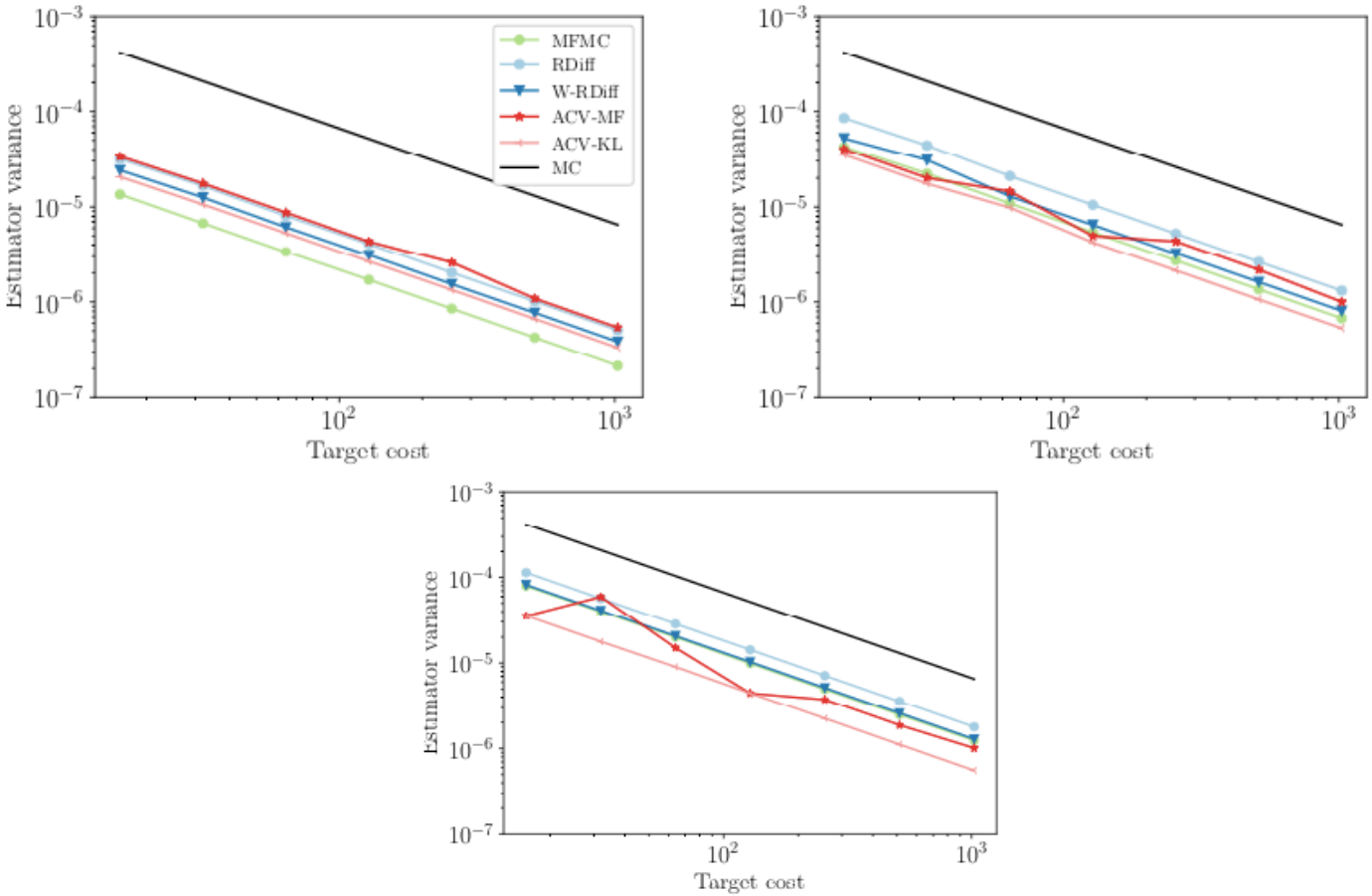
- Elastic wave propagation in two dims. with two materials
- Two numerical schemes, each scheme has different discretizations (10 total simulations)
  - Schemes: High-res second-order scheme, and first-order Godunov scheme
  - Discretizations: 200×200, 100×100, 50×50, 25×25, 10×10

$$q_t + A(q)q_x + B(q)q_y = 0,$$



Robust to aggressive coarsening

1. Left: single-fidelity with discretization hierarchy
2. Right: multi-fidelity with discretization hierarchies
3. Bottom: multi-fidelity with aggressive coarsening



Obervations

Observations:

1. The ACV is a heuristic ansatz, but a useful one as it captures many existing methods
2. Recursive approaches often work really well, general case not always needed
3. Computational gains can be gained by imposing additional structure ***when it is there***
4. Pilot sample are needed for many algorithms, their effect is poorly studied

Alternatives

Type of question we want to answer

- Can the ACV ansatz be ***derived*** from first principles rather than obtained from intuition? (yes)

[Gorodetsky, et. al. *MFNETS: multi-fidelity data-driven networks for Bayesian learning and prediction*, (submitted) 2020]

Alternatives

Type of question we want to answer

- Can the ACV ansatz be ***derived*** from first principles rather than obtained from intuition? (yes)
- Can better, problem-dependent, estimator ansatz be obtained? (probably, ongoing)

[Gorodetsky, et. al. *MFNETS: multi-fidelity data-driven networks for Bayesian learning and prediction*, (submitted) 2020]

Alternatives

Type of question we want to answer

- Can the ACV ansatz be ***derived*** from first principles rather than obtained from intuition? (yes)
- Can better, problem-dependent, estimator ansatz be obtained? (probably, ongoing)

Modeling idea

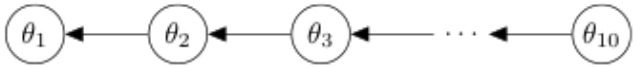
1. Parameterize a set of model relationships, then use first principle inference instead of heuristic approach
2. Will present this from a Bayesian perspective, as it provides a way to incorporate prior knowledge

[Gorodetsky, et. al. *MFNETS: multi-fidelity data-driven networks for Bayesian learning and prediction*, (submitted) 2020]

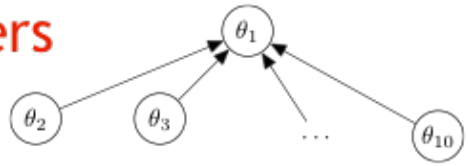
Network models

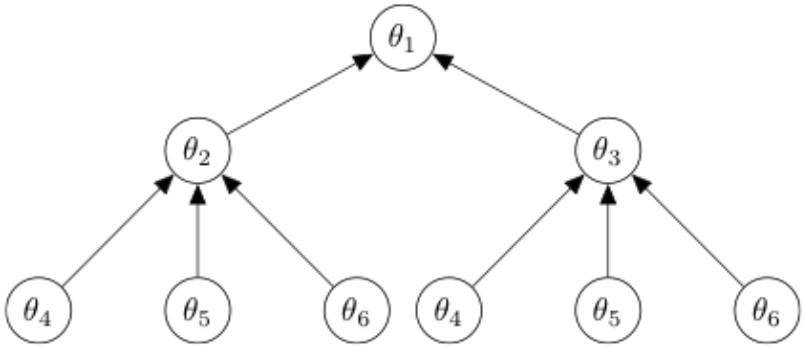
- Model multi-fidelity relationships through conditional independence relations of latent variables
- Fuse multiple models by learning their statistical relationships via network inference

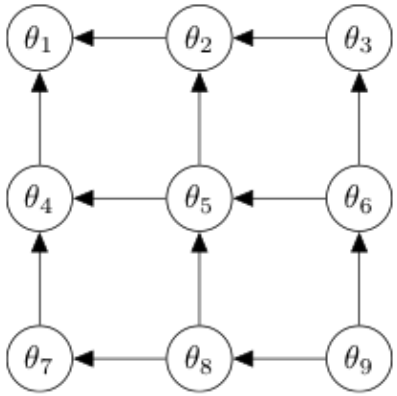
Hierarchy



Low-fidelity peers





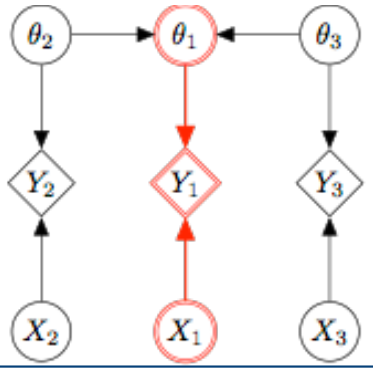


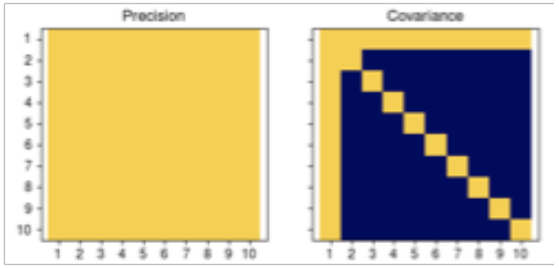
Graphs encode structure, enable scaling to many LV

Peer low-fidelity models

Example: M1 uses a composite turbulence model, M2 and 3 use components

$$p(\theta_1, \theta_2, \theta_3) = p(\theta_3)p(\theta_2)p(\theta_1|\theta_2, \theta_3)$$



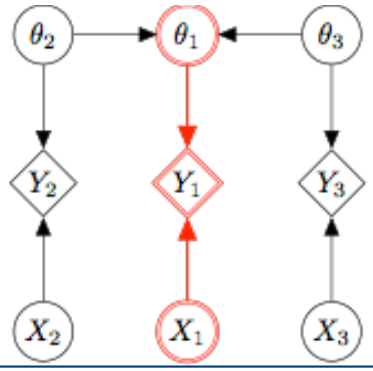


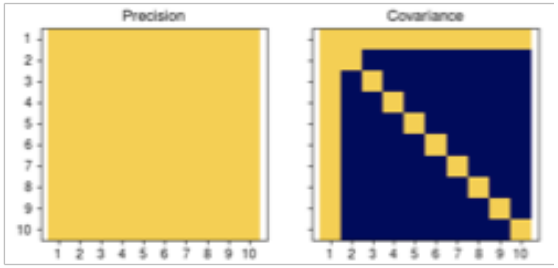
Graphs encode structure, enable scaling to many LV

Peer low-fidelity models

Example: M1 uses a composite turbulence model, M2 and 3 use components

$$p(\theta_1, \theta_2, \theta_3) = p(\theta_3)p(\theta_2)p(\theta_1|\theta_2, \theta_3)$$

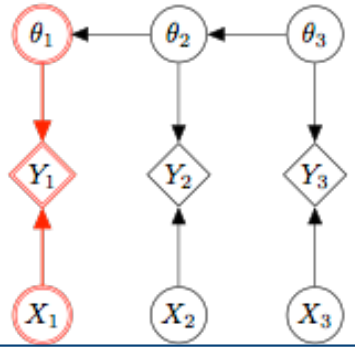


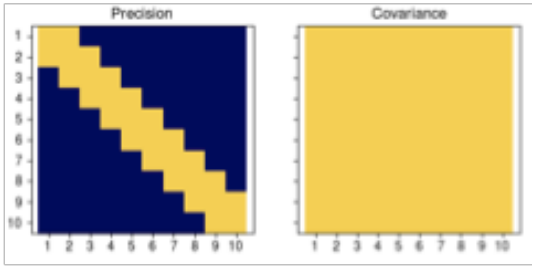


Distinct model hierarchies

Example: refined discretization as in a multilevel scheme

$$p(\theta_1, \theta_2, \theta_3) = p(\theta_3)p(\theta_2|\theta_3)p(\theta_1|\theta_2)$$



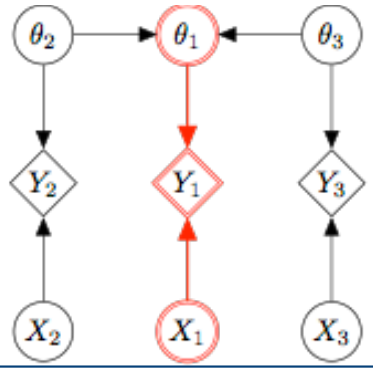


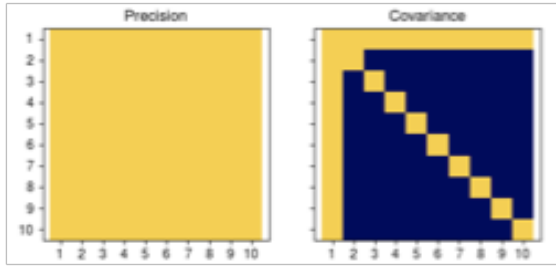
Graphs encode structure, enable scaling to many LV

Peer low-fidelity models

Example: M1 uses a composite turbulence model, M2 and 3 use components

$$p(\theta_1, \theta_2, \theta_3) = p(\theta_3)p(\theta_2)p(\theta_1|\theta_2, \theta_3)$$

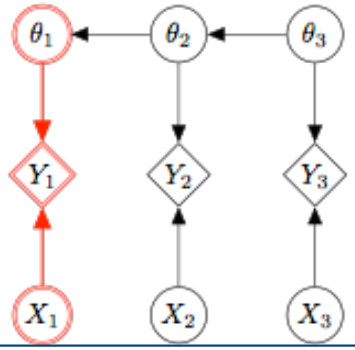


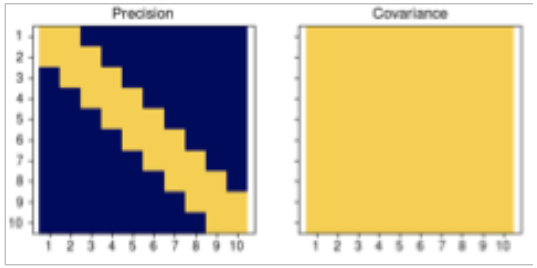


Distinct model hierarchies

Example: refined discretization as in a multilevel scheme

$$p(\theta_1, \theta_2, \theta_3) = p(\theta_3)p(\theta_2|\theta_3)p(\theta_1|\theta_2)$$

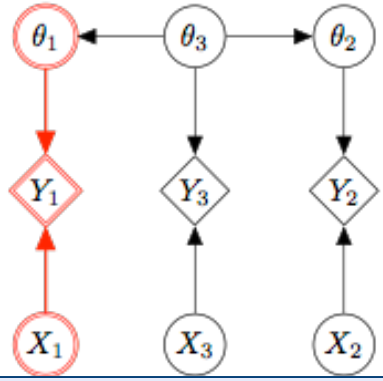


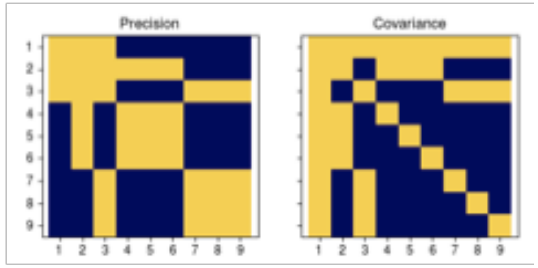


Peer high fidelity models

Example: independent high-fidelity models with an overlapping prediction

$$p(\theta_1, \theta_2, \theta_3) = p(\theta_3)p(\theta_1|\theta_3)p(\theta_2|\theta_3)$$





Bayesian updating and recovering control variates

- Single fidelity:

$$y = \theta + \xi$$

Bayesian updating and recovering control variates

- Single fidelity:

$$y = \theta + \xi$$

  - Prior is $\mathcal{N}(\mu_{prior}, c^2_{prior})$ and likelihood is $\mathcal{N}(\theta, \sigma_1^2)$
  - Posterior mean:

$$\mu_{post} = (1-\nu)\mu_{prior} + \nu\hat{\theta}^{MC}$$

Bayesian updating and recovering control variates

- Single fidelity:

$$y = \theta + \xi$$

  - Prior is $\mathcal{N}(\mu_{prior}, c^2_{prior})$ and likelihood is $\mathcal{N}(\theta, \sigma_1^2)$
  - Posterior mean:

$$\mu_{post} = (1-\nu)\mu_{prior} + \nu\hat{\theta}^{MC}$$

  - Posterior mean is a convex combination of the prior and the MC estimate

Bayesian updating and recovering control variates

- Single fidelity:

$$y = \theta + \xi$$

  - Prior is $\mathcal{N}(\mu_{prior}, c^2_{prior})$ and likelihood is $\mathcal{N}(\theta, \sigma_1^2)$
  - Posterior mean:

$$\mu_{post} = (1-\nu)\mu_{prior} + \nu\hat{\theta}^{MC}$$

  - Posterior mean is a convex combination of the prior and the MC estimate

- Bi-fidelity:

$$y_1 = \theta_1 + \xi, \quad y_2 = \theta_2 + \xi, \text{ and } \theta_1 = a\theta_2 + b + \xi$$

Bayesian updating and recovering control variates

- Single fidelity:

$$y = \theta + \xi$$

  - Prior is $\mathcal{N}(\mu_{prior}, c^2_{prior})$ and likelihood is $\mathcal{N}(\theta, \sigma_1^2)$
  - Posterior mean:

$$\mu_{post} = (1-\nu)\mu_{prior} + \nu\hat{\theta}^{MC}$$

  - Posterior mean is a convex combination of the prior and the MC estimate

- Bi-fidelity:

$$y_1 = \theta_1 + \xi, \quad y_2 = \theta_2 + \xi, \text{ and } \theta_1 = a\theta_2 + b + \xi$$

  - Posterior is convex combination + correction

$$\mu_{1,post} = (1-\nu_2)\mu_{1,prior} + \nu_2\hat{\theta}_1^{MC} + B(\hat{\theta}_2^{MC} - \mu_{2,prior})$$

Bayesian updating and recovering control variates

- Single fidelity:

$$y = \theta + \xi$$

  - Prior is $\mathcal{N}(\mu_{prior}, c^2_{prior})$ and likelihood is $\mathcal{N}(\theta, \sigma_1^2)$
  - Posterior mean:

$$\mu_{post} = (1-\nu)\mu_{prior} + \nu\hat{\theta}^{MC}$$

  - Posterior mean is a convex combination of the prior and the MC estimate

- Bi-fidelity:

$$y_1 = \theta_1 + \xi, \quad y_2 = \theta_2 + \xi, \text{ and } \theta_1 = a\theta_2 + b + \xi$$

  - Posterior is convex combination + correction

$$\mu_{1,post} = (1-\nu_2)\mu_{1,prior} + \nu_2\hat{\theta}_1^{MC} + B(\hat{\theta}_2^{MC} - \mu_{2,prior})$$

  - Compare with CV: recover the CV formula with weight B defined by first principles

Bayesian updating and recovering control variates

- Single fidelity:

$$y = \theta + \xi$$

  - Prior is $\mathcal{N}(\mu_{prior}, c^2_{prior})$ and likelihood is $\mathcal{N}(\theta, \sigma_1^2)$
  - Posterior mean:

$$\mu_{post} = (1-\nu)\mu_{prior} + \nu\hat{\theta}^{MC}$$

  - Posterior mean is a convex combination of the prior and the MC estimate

- Bi-fidelity:

$$y_1 = \theta_1 + \xi, \quad y_2 = \theta_2 + \xi, \text{ and } \theta_1 = a\theta_2 + b + \xi$$

  - Posterior is convex combination + correction

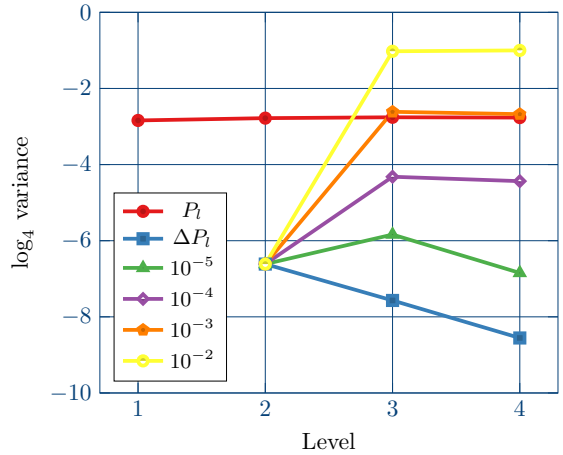
$$\mu_{1,post} = (1-\nu_2)\mu_{1,prior} + \nu_2\hat{\theta}_1^{MC} + B(\hat{\theta}_2^{MC} - \mu_{2,prior})$$

  - Compare with CV: recover the CV formula with weight B defined by first principles

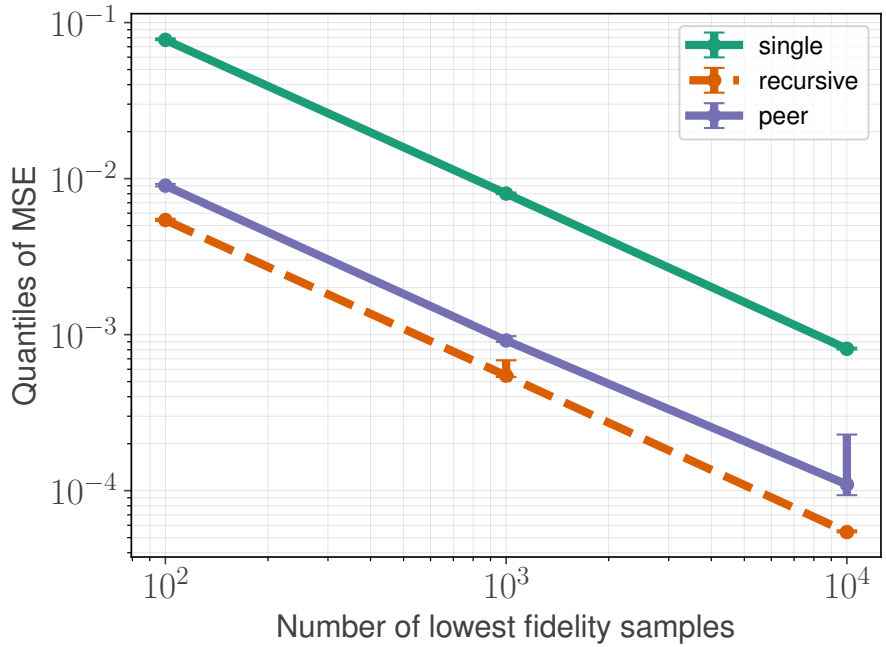
- See paper for theory where we recover CV for the general case

Back to noise robustness example

- Bayesian inference with model selection
- **Common marginal priors on all models**

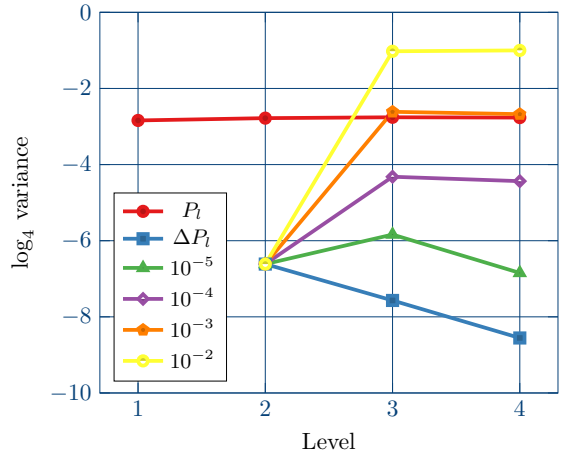


No perturbation

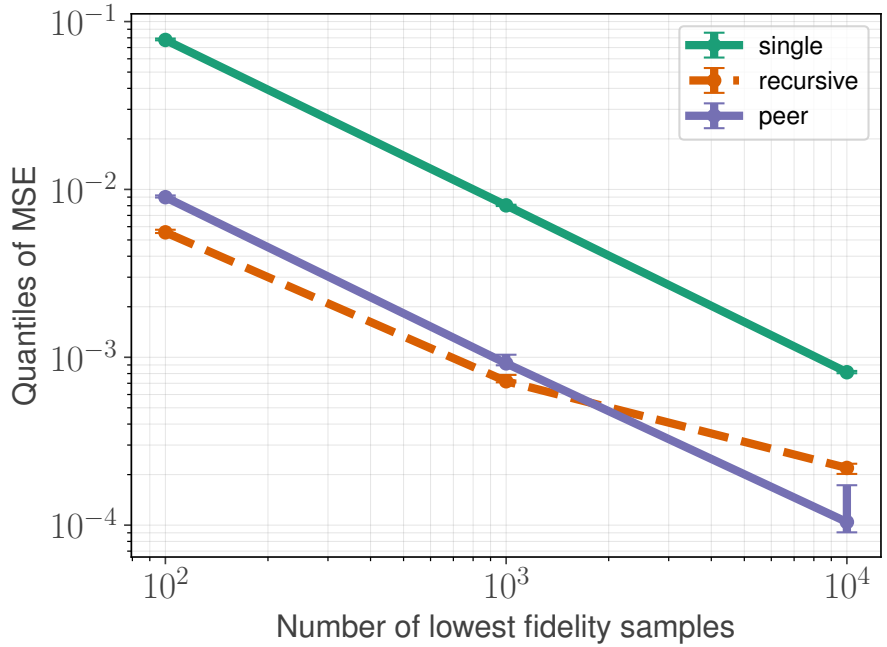


Back to noise robustness example

- Bayesian inference with model selection
- **Common marginal priors on all models**

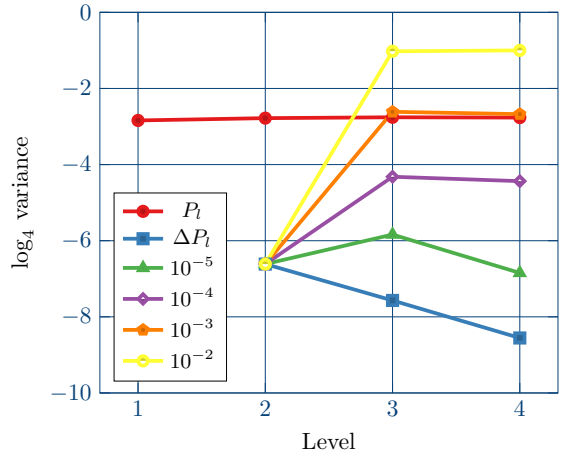


1/100 perturbation

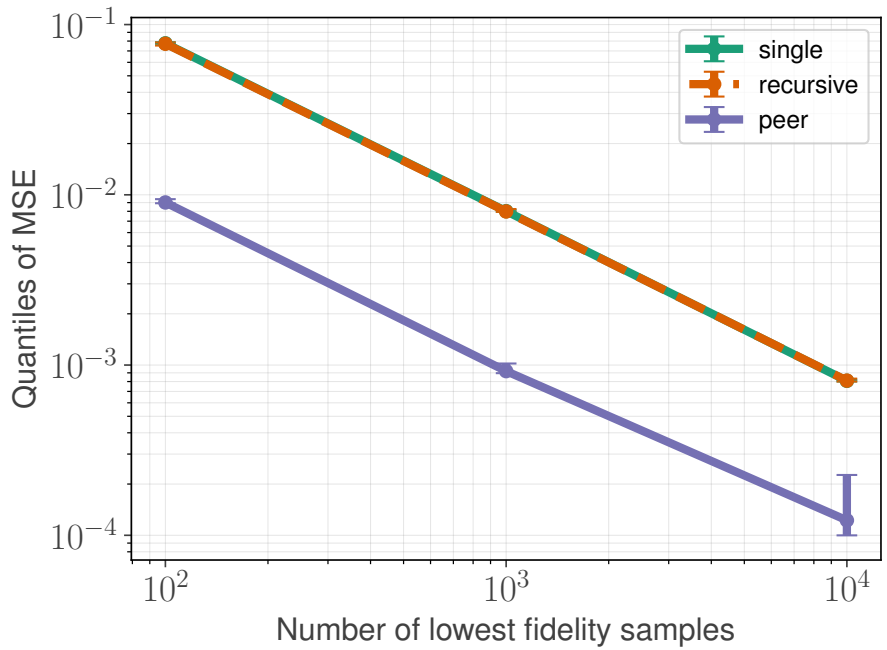


Back to noise robustness example

- Bayesian inference with model selection
- **Common marginal priors on all models**



1/10 perturbation



Summary

- We have developed an approximate control variate framework for multifidelity UQ
  - Works with generic model "fidelities", without ordering based on discretization or correlation
  - Includes recursive difference estimator and MFMC as subsets
  - Enables the development of convergent schemes

Summary

- We have developed an approximate control variate framework for multifidelity UQ
  - Works with generic model "fidelities", without ordering based on discretization or correlation
  - Includes recursive difference estimator and MFMC as subsets
  - Enables the development of convergent schemes
- We have begun to look at the higher-level problem of modeling relationships between general ensembles via probabilistic approaches based on conditional independence relationships

Summary

- We have developed an approximate control variate framework for multifidelity UQ
  - Works with generic model "fidelities", without ordering based on discretization or correlation
  - Includes recursive difference estimator and MFMC as subsets
  - Enables the development of convergent schemes
- We have begun to look at the higher-level problem of modeling relationships between general ensembles via probabilistic approaches based on conditional independence relationships
- Still needed:
  - Adaptation to discretization-based schemes for "bias" reduction (like MLMC, MIMC, and related)
  - Adaptation to surrogate approaches

Summary

- We have developed an approximate control variate framework for multifidelity UQ
  - Works with generic model "fidelities", without ordering based on discretization or correlation
  - Includes recursive difference estimator and MFMC as subsets
  - Enables the development of convergent schemes
- We have begun to look at the higher-level problem of modeling relationships between general ensembles via probabilistic approaches based on conditional independence relationships
- Still needed:
  - Adaptation to discretization-based schemes for "bias" reduction (like MLMC, MIMC, and related)
  - Adaptation to surrogate approaches
- Papers links: see my website

  `https://www.alexgorodetsky.com/publications`

Acknowledgements

This research was funded by the Defense Advanced Research Projects Agency (DARPA) under the EQUiPS program, and the Sandia National Laboratories LDRD program.